Learning the Face Prior for Bayesian Face Recognition
Chaochao Lu, Xiaotang Tang
Dept. of Information Engineering, The Chinese University of Hong Kong, HK.

Introduction
Face recognition mainly consists of two sub-problems: face verification (i.e., to verify whether a pair of face images are from the same person) and face identification (i.e., to recognize the identity of a query face image given a gallery face set). Among the face verification methods, Bayesian face recognition is a representative and successful one.

Classic Bayesian Face Recognition

- The difference between two faces \( x_1 \) and \( x_2 \):
- It classifies \( x \) as intra-personal variations \( D_a \) or extra-personal variations \( D_e \).

- Based on the MAP (Maximum a Posteriori) rule, the similarity measure between \( x_1 \) and \( x_2 \) can be expressed by

\[
s(x_1, x_2) = \log \frac{p(D_a | x_1) p(D_e | x_2)}{p(D_e | x_1) p(D_a | x_2)}
\]

(1)

where both \( p(D_a | x_1) \) and \( p(D_e | x_2) \) are assumed to follow one multivariate Gaussian distribution.

Problems in Classic Bayesian Face Recognition

- Problem 1: It is based on the difference of a given face pair, which discards the discriminative information and reduce the separability.

- Problem 2: The distributions of \( p(D_a | x_1) \) and \( p(D_e | x_2) \) are over-simplified, assuming one multivariate Gaussian distribution can cover large variations in facial poses, illuminations, expressions, aging, occlusions, makeups and hair styles in the real world.

In this paper, we focus on solving Problem 2. (Problem 1 has been addressed in [1], where the joint distribution of \( (x_1, x_2) \) is directly modeled as a Gaussian.)

Key Idea

To overcome Problem 2, we propose a method to automatically learn the conditional distributions of \( (x_1, x_2) \), denoted by \( p(x_1 | x_2) / p(x_2 | x_1) \) and \( p(x_1, x_2) / p(x_1) p(x_2) \).

Our method mainly consists of two steps:

- We exploit the properties of Manifold Relevance Determination (MRD) [2] and extend it to learn the identity subspace for \( (x_1, x_2) \) automatically and accurately.

- Based on the structure of the learned identity subspace, we propose to flexibly estimate Gaussian mixture densities for \( (x_1, x_2) \) with Gaussian process regression [3].

Since the subspace only contains the identity information, the learned density can flexibly reflect the distribution of identities of face pairs \( (x_1, x_2) \) in the observation space.

Preliminary

Gaussian Processes (GPs)

- A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution.

- Infinite set of variables \( \{f(x) : x \in X \} \)

- ARD kernel

\[
h((x_i, x_j)) = (\mathbf{K}(x_i, x_j) + \sigma^2 \mathbf{I})^{-1} \mathbf{K}(x_i, x_j)
\]

(8)


Handling Large Poses

Verify the validity of learning identity subspaces in our approach. (b) To verify the validity of learning the distributions of identity in our approach.

- To verify the relationship between the number of images for each individual and the performance of our approach.

- Comparison with other Bayesian face methods.